Vision Research 51 (2011) 546–552

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Oculomotor capture during real-world scene viewing depends on cognitive load

Michi Matsukura [a,*], James R. Brockmole [b], Walter R. Boot [c], John M. Henderson [d]

[a] University of Iowa, Department of Psychology, 11 Seashore Hall E, Iowa City, IA 52242, USA
[b] University of Notre Dame, Department of Psychology, Haggar Hall, Notre Dame, IN 46556, USA
[c] Florida State University, Department of Psychology, 1107 W Call Street, Tallahassee, FL 32306, USA
[d] University of South Carolina, Department of Psychology and McCausland Center for Brain Imaging, Columbia, SC, 29208, USA

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 20 January 2011
Available online 15 February 2011

Keywords:
Gaze control
Oculomotor capture
Real-world scene viewing
Attention capture

ABSTRACT

It has been claimed that gaze control during scene viewing is largely governed by stimulus-driven, bottom-up selection mechanisms. Recent research, however, has strongly suggested that observers' top-down control plays a dominant role in attentional prioritization in scenes. A notable exception to this strong top-down control is oculomotor capture, where visual transients in a scene draw the eyes. One way to test whether oculomotor capture during scene viewing is independent of an observer's top-down goal setting is to reduce observers' cognitive resource availability. In the present study, we examined whether increasing observers' cognitive load influences the frequency and speed of oculomotor capture during scene viewing. In Experiment 1, we tested whether increasing observers' cognitive load modulates the degree of oculomotor capture by a new object suddenly appeared in a scene. Similarly, in Experiment 2, we tested whether increasing observers' cognitive load modulates the degree of oculomotor capture by an object's color change. In both experiments, the degree of oculomotor capture decreased as observers' cognitive resources were reduced. These results suggest that oculomotor capture during scene viewing is dependent on observers' top-down selection mechanisms.

1. Introduction

Given the complexity of the visual world, observers must select a subset of possible visual inputs for processing. Which particular inputs (locations or objects) are selected to receive processing priority is determined, in part, by an observer's behavioral goals (e.g., Henderson & Hollingworth, 1999; Yarbus, 1967). However, some visual events can attract attention when they have little or no relationship to the observer's intended behavior (e.g., Irwin, Colcombe, Kramer, & Hahn, 2000; Theeuwes, 1994; Theeuwes, Kramer, Hahn, & Irwin, 1998; Yantis & Jonides, 1984). In these situations, attention is referred to be *captured*. Studies using relatively simple displays of geometric shapes and letters have demonstrated that various types of unique and novel stimuli attract both covert attention and gaze, the most reliable of these being the appearance of a new object (e.g., Boot, Kramer, & Peterson, 2005b; Irwin et al., 2000; Theeuwes, 1994; Theeuwes et al., 1998; Yantis & Jonides, 1984).

A series of recent studies has extended investigation of overt attention capture (also referred to as *oculomotor capture*) to the appearance of new objects in real-world scenes (Brockmole & Henderson, 2005a, 2005b, 2008; Matsukura, Brockmole, & Henderson, 2009). In these studies, observers viewed a series of scenes under the guise of preparing for a later memory test (which was not actually given). During viewing, a new object was suddenly added to the scene during a fixation so that it was not masked by saccadic suppression. The extent to which these changes captured attention was measured by observing the propensity for observers' eyes to be directed to the regions in which these onsets occurred (cf., Irwin et al., 2000; Theeuwes et al., 1998). While the chance rate of viewing objects in scenes without onsets was approximately 10%, when onsets were present in scenes, roughly 60% of the first eye movements following the onsets were allocated to the new objects.[1] Thus, onsets in scenes attract attention and gaze quickly and reliably. Moreover, these capture effects have been shown to be independent of task instruction (Brockmole & Henderson, 2005a) and semantic identity of the onsets (Brockmole & Henderson, 2008).

The oculomotor capture findings described above have been interpreted as evidence that gaze control is sometimes driven by stimulus-based selection mechanisms. Similar conclusions have also been drawn from studies linking local image statistics (e.g. Baddeley & Tatler, 2006; Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Mannan, Ruddock, & Wooding, 1995, 1996; Mannan, Ruddock, & Wooding, 1997; Parkhurst & Niebur, 2003;

* Corresponding author. Address: Department of Psychology, 11 Seashore Hall E, University of Iowa, Iowa City, IA 52242-1407, United States. Fax: +1 319 512 5958.
E-mail address: michi-matsukura@uiowa.edu (M. Matsukura).

[1] In the present study, we use "a sudden appearance of a new object" and "an onset" in the context of scene viewing interchangeably.

0042-6989/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2011.01.014

Reinagel & Zador, 1999) and visual salience (e.g. Itti & Koch, 2000; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002; Rosenholtz, 1999) to fixation placement. However, the idea that such low-level image properties can contribute to gaze control independently of an observer's top-down knowledge has also received a wide range of criticisms (Foulsham & Underwood, 2007; Henderson, 2003; Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson, Malcolm, & Schandl, 2009; Pelz & Canosa, 2001; Torralba, Oliva, Castelhano, & Henderson, 2006; Turano, Geruschat, & Baker, 2003). As a result, some researchers have argued that scene-based oculomotor capture effects serve as the best evidence for a stimulus-driven selection mechanism that supersedes observers' cognitive control of gaze (e.g., Henderson et al., 2007). The purpose of the present study was to directly test this hypothesis.

To examine whether oculomotor capture during scene viewing is indeed independent of cognitive control, we employed a dual-task paradigm that has previously been used to address the stimulus-driven nature of covert capture by onsets. The logic behind this paradigm is the following: If attention capture is truly independent of observers' top-down control mechanisms, then stimulus-driven processes should be impervious to manipulations of observers' cognitive load. For example, Boot, Brockmole, and Simons (2005a) had one group of observers search for a target letter in a letter array. During this search, an additional irrelevant letter was suddenly added to the array (onset). A second group of observers performed the same search task while also engaged in a demanding concurrent auditory counting task. While the search-only group exhibited robust capture, onsets failed to influence search for those in the dual-task group. Based on these results, Boot et al. concluded that attention capture cannot be purely stimulus-driven, given that it is modulated by cognitive load.

In the present study, we examined whether oculomotor capture during scene viewing is similarly modulated while observers perform a cognitively demanding secondary task. If capture of new objects in real-world scenes is indeed stimulus-driven, then the attentional priority given to onsets should not be affected by whether an observer is performing a concurrent secondary task or not. By contrast, if oculomotor capture arises from similar mechanisms to covert attention capture (Hunt, von Mühlenen, & Kingstone, 2007), then observers' engagement in an attention-demanding concurrent task should modulate the probability that oculomotor capture occurs. This latter result would suggest that oculomotor capture during real-world scene viewing is not independent of observers' top-down selection mechanisms.

2. Experiment 1

Experiment 1 investigated whether variations in cognitive load modulate the degree of oculomotor capture generated by the sudden appearance of a new object during real-world scene viewing. We combined the scene-based oculomotor capture paradigm introduced by Brockmole and Henderson (2005a) and the dual-task capture paradigm developed by (Boot et al. 2005a, also see a similar manipulation used in Lavie & de Fockert, 2005).

2.1. Method

2.1.1. Participants

Twenty-four undergraduates with normal or corrected-to-normal vision were paid for their participation in a single 30-min experimental session.

2.1.2. Visual stimuli

Stimuli consisted of full-color photographs of 30 real-world scenes. These were the same stimuli described in Matsukura et al. (2009). Initially, two photographs of each scene were taken, differing only in the presence or absence of a single *critical object* (Fig. 1, Top Panels). Photographs were digitally edited to eliminate minor differences in shadow and spatial displacement between each shot. Local luminance was closely approximated in each scene version. Photographs were displayed at a resolution of 800 × 600 pixels in 24-bit color and subtended 37° horizontally and 27.5° vertically at a viewing distance of 81 cm.

2.1.3. Auditory stimuli

Strings of 10 single-digit numbers were articulated by a digitized voice at a rate of 2 digits/s for 5 s. Digit strings were randomly generated for each trial with the constraint that they included either two or three sequential digit repetitions. For example, the string 1, 9, 4, 4, 5, 8, 3, 3, 6, had two sequential repetitions (the 4's and 3's). Observers were told that up to four repetitions could occur in order to elicit continued attention to the auditory stream after three repetitions.

2.1.4. Apparatus

Visual stimuli were presented on a 21-in. CRT monitor with a screen refresh rate of 120 Hz. Throughout each trial, the spatial position of each observer's right eye was sampled at a rate of 1000 Hz by a tower-mounted EyeLink 2 K eye-tracking system (SR research, Inc.) running in pupil and corneal-reflection mode, resulting in an average spatial accuracy of 0.15°. An eye movement was classified as a saccade if its amplitude exceeded .2° and either (a) its velocity exceeded 30°/s or (b) its acceleration exceeded 9500°/s. Chin and forehead rests stabilized head position and kept viewing distance constant. Auditory stimuli were presented via stereo speakers placed directly below the visual display.

2.1.5. Design and procedure

Observers were randomly assigned to one of two between-subjects conditions. In the *onset condition*, a critical object was added to each scene during viewing (details below). In the *control condition*, the same critical object was visible throughout the trial. The control condition allowed us to determine the baseline rate at which the onset object was fixated when it was not suddenly added during viewing. Whether in the control or onset condition, all the observers viewed the scenes under two task loads. In the *single-task* condition, the observers viewed each scene while ignoring a concurrent auditory stimulus. In the *dual-task* condition, the

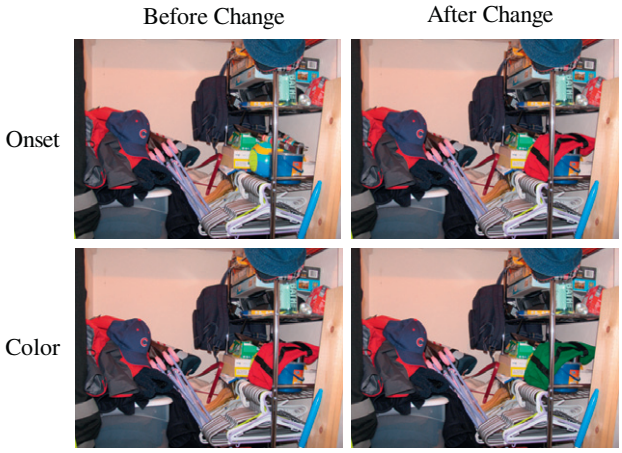

Fig. 1. An example scene used in the current study, for both before (left panels) and after (right panels) the scene change. Top: Onset (Experiment 1), Bottom: Color Change (Experiment 2). To view this figure in color, please see the online version of this article.

observers viewed each scene while counting the number of sequential repetitions within the auditory number string. For each observer, 15 scenes were randomly selected to be included in the single-task condition while another 15 scenes were presented in the dual-task condition. Single-task and dual-task trials were blocked and the order of these blocks was counterbalanced across the observers. In all cases, the observers' primary task was to memorize the scene in preparation for a subsequent memory test which was to be administered after all scenes were studied.

The observers began the experimental session by completing a calibration routine that mapped the output of the eye tracker onto display position. Calibration was constantly monitored throughout the experiment and was adjusted when necessary. The observers began each trial by fixating a dot in the center of the display. After pressing a button to initiate the trial, a photograph and voice string were presented for 5 s (i.e., the auditory stream started when a scene was presented, and concluded when this scene was removed). In the onset condition, an object was added while an observer was studying a scene by seamlessly switching the photograph presented on the display with its associated counterpart that contained the additional object. These onsets were yoked to the first saccadic eye movement that occurred after 3 s had elapsed from the beginning of the trial. Specifically, onsets were executed 100 ms after the start of this saccade. This 100-ms delay was long enough to allow the saccade to terminate but short enough that a subsequent saccade was unlikely to be launched before the onset. Thus, the eyes were stable when the onsets occurred (see Brockmole & Henderson, 2005a, 2005b; Brockmole & Henderson, 2008; Matsukura et al., 2009). In order to avoid head movements associated with speaking, in the dual-task condition, the observers signed their secondary task response with their fingers at the conclusion of each trial and this was recorded by the experimenter.

After viewing all 30 scenes (15 single-task scenes, 15 dual-task scenes), each observer completed a memory test. Stimuli consisted of color photographs of 60 real-world scenes. Thirty of these scenes were the post-change pictures presented during the study sessions (scenes presented in the single-task and dual-task scenes). The other 30 scenes were new scenes that were not previously shown to observers but were similar in spatial scale, structure and content (control scenes). The observers made an un-speeded response using a keypad to indicate whether or not each picture was presented during the initial scene viewing period.

2.2. Results and discussion

2.2.1. Preliminary analyses

Examination of the eye movement record indicated that new objects successfully appeared during a fixation on 97% of trials in the single-task condition and on 92% of trials in the dual-task condition. All remaining trials were excluded from the reported analyses. Mean accuracy for the auditory task was 83% for the onset condition and 90% for the control condition, $F\ (1, 22) = 1.60$, $p = .22$. In terms of subsequent memory performance, the observers accurately recognized 92% of the scenes presented in the single-task condition and 63% of the scenes presented in the dual-task condition, $F\ (1, 11) = 23.64$, $p < .001$.[2] No significant accuracy difference was observed between control scenes (94%) and single-task scenes (92%), $F\ (1, 11) = 0.10$, $p = .75$. These results verify that the secondary task placed substantial cognitive load on the observers. Our main question of interest was whether or not this load modulated the degree of onset-induced oculomotor capture.

[2] The analysis of A' yielded the same pattern of the results as did the analysis of percent correct for both Experiments 1 and 2.

2.2.2. Onset-induced oculomotor capture

For each load condition (single-task vs. dual-task), we determined how often and how quickly the onset was fixated (see Brockmole & Henderson, 2005a).

2.2.2.1. Frequency of capture. For each scene, a region of interest was defined by the smallest imaginary rectangle that could surround the critical object. Fixations were sorted based on whether they fell within or outside these regions of interest. We restricted our analysis to the first four fixations following the onset. We denote these as *ordinal fixation positions* 1, 2, 3, and 4, respectively. Fixation 1 corresponds to the termination of the first saccade launched after the onset. Therefore, it is the first fixation that could be influenced by the onset. If onsets capture gaze, then observers' eyes should be directed to the location of the onset with greater-than-chance probability. This chance level was obtained from the control condition where, on average, 8% of fixations were localized on the critical object (this *baseline rate of viewing* did not significantly differ between the single-task and dual-task conditions, $t\ (11) = .06$, $p = .55$). If an onset draws attention, then the fixation probability should exceed the baseline rate. Indeed, 95% confidence intervals indicated that onsets were fixated more frequently than the baseline rate of viewing at all four ordinal fixation positions for both the single-task and dual-task conditions (see Fig. 2, Top Panel).

A 2 (load) × 4 (ordinal fixation position) repeated-measures analysis of variance (ANOVA) was conducted to determine whether the frequency of fixating the onset varied as a function of load (single-task vs. dual-task) and ordinal fixation position (Fixations 1–4). The observers fixated the onset more often when they were engaged in the viewing task only (61% of trials) than when they were engaged in both the viewing and auditory tasks (36% of trials), $F\ (1, 11) = 24.77$, $p < .0001$. Onsets were not fixated equally at all ordinal fixation positions, which led to a significant main effect of ordinal fixation position, $F\ (3, 33) = 7.54$, $p < .003$. After peaking at Fixation 2, fixations on the onset in the single-task condition rapidly declined. In contrast, the probability of fixating the onset

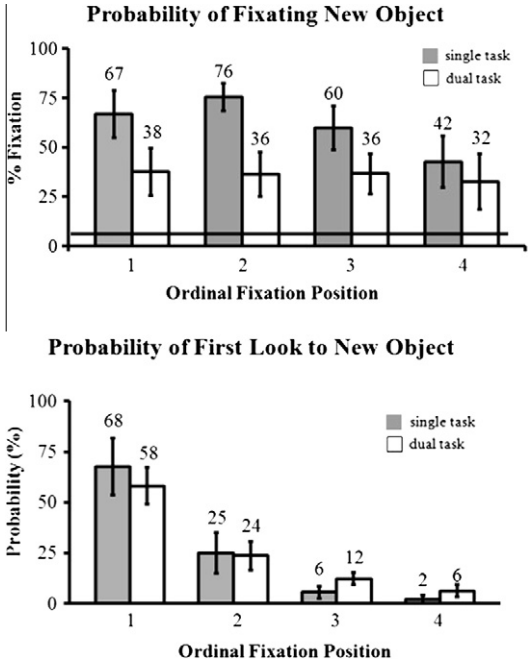

Fig. 2. Results, Experiment 1. Top: The mean probability of fixating the onset as a function of load (single-task vs. dual-task) and ordinal fixation position (Fixations 1–4). The solid line illustrates the baseline rate of viewing (chance). Bottom: The probability with which the first look to the onset occurred at each of the first four fixations after the onset.

remained stable across all four ordinal fixation positions in the dual-task condition. This difference led to a significant interaction of load and ordinal fixation position, $F(3, 33) = 2.82$, $p < .05$. In fact, when the data from the single-task and dual-task conditions were analyzed separately, the main effect of ordinal fixation position was significant for the single-task condition, $F(3, 33) = 9.75$, $p < .0001$, but not for the dual-task condition, $F(3, 33) = .21$, $p = .89$. Planned pair-wise comparisons confirmed that onsets were fixated significantly more often in the single-condition than in the dual-task condition at Fixations 1–3, $t(11) = 3.00$, $p < .01$, $t(11) = 6.91$, $p < .0001$, $t(11) = 3.82$, $p < .0002$, respectively, but not at Fixation 4, $t(11) = 1.03$, $p = .33$. These results indicate that oculomotor capture is less likely to occur under higher cognitive load.

2.2.2.2. Speed of capture. While the analysis of gaze location across the first four post-onset fixations provides a measure of the frequency of capture, the combination of first looks to and re-fixations on new objects prevents us from obtaining a clean picture of the speed with which onsets were prioritized. To obtain a clearer measure of speed, we computed the number of times the first look to the onset occurred at each ordinal fixation position.

A 2 (load) × 4 (ordinal fixation position) repeated-measures ANOVA was conducted (see Fig. 2, Bottom Panel). To avoid issues of multi-collinearity introduced by expressing the number of first looks to scene changes at each ordinal fixation position as a conditional probability, we performed the ANOVA on the raw number of times that the first look occurred at each fixation position (see Brockmole & Henderson, 2005a, for this method). Mirroring the frequency of capture analysis (Fig. 2, Top Panel), more first looks to the onset were observed in the single-task condition than the dual-task condition, $F(1, 11) = 20.77$, $p < .001$. In terms of ordinal fixation position, first looks to the onset occurred most frequently at Fixation 1, followed by a rapid decline across Fixations 2–4, which led to a significant main effect of ordinal fixation position, $F(3, 33) = 42.36$, $p < .0001$. Critically, the significant interaction between load and ordinal fixation position indicated that first looks to the onset in the single-task and dual-task conditions were not similarly distributed across fixation positions, $F(3, 33) = 42.36$, $p < .0001$. As is apparent in Fig. 2 (Bottom Panel), more first looks to the onset occurred earlier during viewing in the single-task condition compared to the dual-task condition. In fact, the effect of ordinal fixation position was significant for both the single-task, $F(3, 33) = 25.97$, $p < .0001$, and the dual-task conditions, $F(3, 33) = 14.53$, $p < .0001$. Planned pair-wise comparisons confirmed that significantly more first looks were made to the onsets in the single-task condition than in the dual-task condition at Fixation 1, $t(11) = 2.97$, $p < .01$, but not at Fixations 2–4, $t(11) = 1.53$, $p = .15$, $t(11) = .71$, $p = .49$, $t(11) = 1.39$, $p = .19$, respectively. Consistent with the frequency analysis above, the speed analysis results also indicate that oculomotor capture slows down under higher cognitive load.

2.2.2.3. Summary. The results of Experiment 1 indicate that both the likelihood and speed of oculomotor capture in the face of sudden onsets are reduced in the dual-task condition. These results parallel the pattern observed by Boot et al. (2005a) in a covert capture paradigm that involved arrays of letters, and it presents a strong challenge to the hypothesis that oculomotor capture in real-world scenes is encapsulated from observers' higher cognitive resources. In Experiment 2, we seek converging evidence for this conclusion using color-induced oculomotor capture.

3. Experiment 2

Regardless of whether it is covert or overt, attention capture can be driven by object properties (features) other than onsets. For example, it has been reported that object surface feature such as color can induce attention capture. Task-irrelevant color singletons (Irwin et al., 2000; Theeuwes, 1994) or changes to an object's color (Matsukura et al., 2009) can attract attention. For instance, in Matsukura et al. (2009), the color of an object in a real-world scene was abruptly switched while observers were viewing each scene. Although these color changes were less effective attractors of attention than onsets, they attracted 35–40% of the eye movements launched immediately following the color change (this rate was four times higher than the baseline rate of viewing). The purpose of Experiment 2 was to determine whether cognitive load also influences the degree of color-induced capture.

3.1. Method

The method of Experiment 2 was identical with that of Experiment 1 except for the following. Rather than introducing a new object, an existing object in a scene changed color (Fig. 1, Bottom Panel). These color alterations were achieved within CIE L∗ a∗ b∗ color space while holding luminance constant. Additional details are provided in Matsukura et al. (2009). Twelve new observers participated in this *color-change* condition. The baseline condition from Experiment 1 was used as the control condition in Experiment 2.

3.2. Results and discussion

3.2.1. Preliminary analyses

Preliminary analyses were consistent with Experiment 1. Critical objects successfully changed color during a fixation on 94% of trials in both the single-task and dual-task conditions (remaining trials were excluded from the analyses). Mean accuracy for the auditory task was 89% for the color-change condition and 90% for the control condition, $F(1, 22) = .03$, $p = .87$. In terms of subsequent memory test performance, the observers accurately recognized 98% of the scenes presented in the single-task condition and 74% of the scenes presented in the dual-task condition, $F(1, 11) = 23.64$, $p < .001$. Unlike Experiment 1, the observers' recognition accuracy was higher for scenes presented during the single-task condition (98%) than control scenes (91%), $F(1, 11) = 7.65$, $p < .01$. This difference is likely to derive from the observers' prior experience with the single-task scenes during the viewing task.

3.2.2. Color-induced oculomotor capture

3.2.2.1. Frequency of capture. Ninety-five percent confidence intervals indicated that color changes in both single-task and dual-task conditions were fixated more frequently than the baseline rate of viewing at all four ordinal fixation positions (Fig. 3, Top Panel). A 2 (load) × 4 (ordinal fixation position) repeated-measures ANOVA indicated that the observers fixated the color change more often when they were engaged in the viewing task only (43%) compared to when they were engaged in both the viewing and auditory tasks (27%), $F(1, 11) = 11.76$, $p < .006$. Once again, color changes were not fixated equally at all ordinal fixation positions, which led to a significant main effect of ordinal fixation position, $F(3, 33) = 5.98$, $p < .005$, with viewing peaking at Fixation 2.

More frequent fixations on color changes in the single-task condition than in the dual-task condition across the first three fixation positions failed to produce a significant interaction of load and ordinal fixation position, $F(3, 33) = 2.25$, $p = .1$. The effect of ordinal fixation position was significant in the single-task condition, $F(3, 33) = 9.66$, $p < .0001$, but not in the dual-task condition, $F(3, 33) = .94$, $p = 43$. However, planned pair-wise comparisons revealed that significantly more fixations was made in the single-task condition than in the dual-task condition at Fixations 1–3, $t(11) = 3.56$, $p < .004$, $t(11) = 4.02$, $p < .002$, $t(11) = 2.85$, $p < .02$,

respectively, but not at Fixation 4, t (11) = .67, p = .52. As observed in onset-induced oculomotor capture (Experiment 1), these results indicate that color-induced oculomotor capture is less likely to occur under higher cognitive load.

3.2.2.2. Speed of capture. As in Experiment 1, a 2 (load) × 4 (ordinal fixation position) repeated-measures ANOVA was conducted to compare the number of first looks to the color change at each ordinal fixation position (Fig. 3, Bottom Panel). More first looks to the color change were observed in the single-task condition relative to the dual-task condition, F (1, 11) = 17.82, p < .001, and a sharp drop was observed from Fixations 1–4 for both single-task and dual-task conditions, F (3, 33) = 27.79, p < .0001. However, the significant interaction between load and ordinal fixation position indicated that first looks to the color change in the single-task and dual-task conditions were not similarly distributed across fixation positions, F (3, 33) = 3.92, p < .02. As it is apparent in Fig. 3 (Bottom Panel), more first looks to color changes occurred earlier during scene viewing in the single-task condition compared to the dual-task condition. The effect of ordinal fixation position was significant for both the single-task condition, F (3, 33) = 22.37, p < .0001, and the dual-task condition, F (3, 33) = 12.12, p < .0001. Planned pairwise comparisons confirmed that significantly more first looks were directed to color changes in the single-task condition than in the dual-task condition at Fixation 1, t (11) = 3.2, p < .008, but not at Fixations 2–4, t (11) = .8, p = .44, t (11) = 1.00, p = .33, t (11) = −.32, p = .75, respectively. While the frequency analysis above failed to produce a significant interaction of load and ordinal fixation position, the speed analysis showed a significant interaction of load and ordinal fixation position. This pattern suggests that the observers re-fixated the color change more often than the onset; however, as it will be reported in the later between-experiments analysis, this difference did not reach significance. These results indicate that, as in Experiment 1, color-induced oculomotor capture occurs slower under higher cognitive load.

3.2.2.3. Summary. The results of Experiment 2 are consistent with those of Experiment 1. Both the likelihood and speed of oculomotor capture in the face of sudden color changes were reduced in the dual-task condition. These results provide strong converging evidence that oculomotor capture in real-world scenes is not immune to observers' cognitive load.

3.2.3. Onset-Induced vs. color-induced oculomotor capture

To obtain a clearer picture of likelihood and speed of oculomotor capture caused by different types of visual events (i.e., a sudden appearance of a new object vs. an abrupt color change of the existing object in a scene), we conducted a mixed-model ANOVA that contrasted the patterns of results obtained in Experiments 1 and 2.

3.2.3.1. Frequency of capture. A mixed-model ANOVA with within-subjects factors of load and ordinal fixation position and a between-subjects factor of change type (onset vs. color change) was conducted to determine whether the frequency of fixating the critical object varied as a function of load (single-task vs. dual-task), ordinal fixation position (Fixations 1–4), and change type (onset vs. color change). The observers fixated the critical object more often when they were engaged in the viewing task only than when they were engaged in both the viewing and auditory tasks, F (1, 22) = 35.81, p < .0001. Replicating Matsukura et al. (2009), the observers fixated new objects more frequently than color changes, F (1, 22) = 8.8, p < .007. Scene changes were not fixated equally at all ordinal fixation positions, which led to a significant main effect of ordinal fixation position, F (3, 66) = 13.17, p < .0001. The critical object was fixated more often during Fixation 2 than any other fixation position. Fixation probability in the single-task condition rapidly declined while it did not in the dual-task condition, leading to a significant interaction of load and ordinal fixation position, F (3, 66) = 4.90, p < .004. However, this interaction between load and ordinal fixation position did not differ across different change types, F (3, 66) = .37, p = .77. These results indicate that, regardless of change type, oculomotor capture is less likely to occur when less cognitive resources are available.

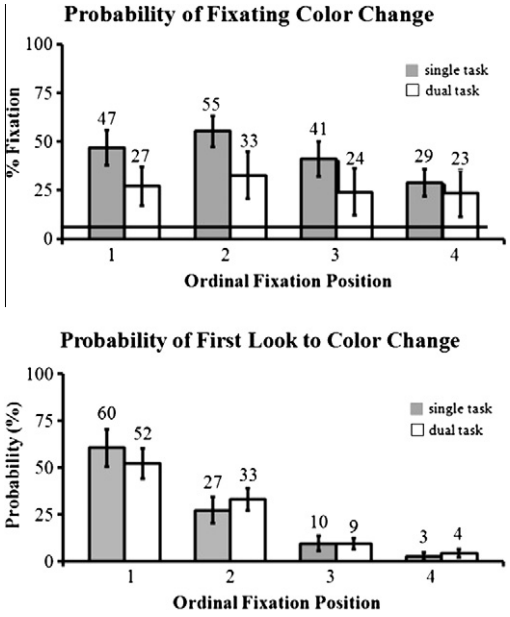

Fig. 3. Results, Experiment 2. Top: The mean probability of fixating color change as a function of load (single-task vs. dual-task) and ordinal fixation position (Fixations 1–4). The solid line illustrates the baseline rate of viewing (chance). Bottom: The probability with which the first look to color change occurred at each of the first four fixations after the onset.

3.2.3.2. Speed of capture. The probability of making a first look to the critical object was also higher in the single-task condition than in the dual-task condition, F (1, 22) = 33.37, p < .0001. Consistent with the frequency analysis above, the appearance of new objects attracted observers' first fixations more often than sudden color changes, F (1, 22) = 8.69, p < .007. These first looks to the critical change occurred significantly faster in the single-task condition than in the dual-task condition, which produced a significant interaction of load and ordinal fixation position, F (3, 66) = 8.38, p < .0001. Because this pattern of faster prioritization in the single-task than in the dual-task condition was consistent across the onset and color-change conditions, the three-way interaction of load, ordinal fixation position and change type did not reach significance, F (3, 66) = .76, p = .52. These results indicate that the reduced cognitive resources retard the speed of oculomotor capture regardless of whether a scene change involved the sudden appearance of the new object or alternation of the existing object's color.

4. General discussion

Gaze control during real-world scene viewing is influenced by both stimulus-driven and cognitive factors (see Henderson, 2007, for a review). Recently, a great deal of research has been conducted to investigate the extent to which stimulus-driven mechanisms influence gaze control independently of observers' knowledge and expectations. However, studies of local image statistics and visual salience have been equivocal at best (e.g. Foulsham & Underwood, 2007; Henderson et al., 2007, 2009; Pelz & Canosa, 2001;

Torralba et al., 2006; Turano et al., 2003), leaving oculomotor capture as the best candidate to examine if a purely bottom-up selection process can override observers' top-down/cognitive intension. In fact, some researchers have suggested that oculomotor capture may represent a case where stimulus-based factors have priority over cognitive factors in controlling fixation placement within scenes (Henderson et al., 2007). The primary purpose of the present study was to examine this hypothesis.

We investigated whether onset-induced and color-induced oculomotor capture during real-world scene viewing is automatic using a dual-task paradigm that has been previously employed in covert attention capture paradigms (e.g., Boot et al., 2005a; Lavie & de Fockert, 2005). In two experiments, we demonstrated that increasing observers' cognitive load during a scene viewing task reduced the frequency and speed of oculomotor capture by both onsets and color changes. These results suggest that even oculomotor capture, a type of gaze behavior that would appear to be a good candidate for complete bottom-up control, is modulated by top-down control. The general conclusion seems to be that, during real-world scene viewing, there is no mechanism component of gaze control that is completely stimulus-driven.

An interesting contrast can be drawn between the results of Experiment 2 and prior research on color singletons. Both Boot et al. (2005a) and Lavie and de Fockert (2005) demonstrated that cognitive load increases capture induced by a color singleton. However, in our Experiment 2, we demonstrated reduced capture by color changes under the dual-task load. At first glance, diverging effects of cognitive load on a color-based distractor may seem incongruous; however, the observed difference can be explained by drawing a distinction between transient and sustained distracting events. Boot et al. (2005a) developed this distinction to account for why cognitive load decreases onset-induced capture but increases color singleton-induced capture (also see Lavie and de Fockert (2005) for a related argument). In a homogenous search array, once a new object is added, the new object does not remain visually unique for an extended period of time (i.e. it is a transient event). In contrast, a color singleton remains distinct from other items for an extended period of time in the homogenous search display (i.e., it is a sustained event). In complex real-world scenes, it is unlikely that any color change results in a color singleton. Because neither an onset nor a color change was visually unique relative to its surroundings over time, both types of scene change can be considered to be transient, and these changes may be more likely to go unnoticed under higher cognitive load.

The current findings can also be linked to other studies that examined the nature of attention capture with dual-task manipulations. We employed a secondary auditory task that did not share a sensory modality with the primary scene viewing task because we were explicitly interested in how competition for general cognitive resources influences oculomotor capture rather than whether specific content (e.g., object features) held in memory affects visual attention. For example, by using a task-irrelevant color singleton search task (Olivers, 2009; Olivers, Meijer, & Theeuwes, 2006) demonstrated that search latency increased when the singleton distractor matched memory content (but see Woodman & Luck, 2007; also see Han & Kim, 2009 for the effect of perceptual difficulty and time course of cognitive control), and this interference was strong only when the content of memory was inherently visual. While Olivers et al.'s study used an attention capture paradigm to investigate whether visual attention and visual working memory shared the same content representations, we asked whether a scene change could still be prioritized when less cognitive resources were available. It remains an interesting question as to whether content specific memory effects influence oculomotor capture in real-world scenes.

Having acknowledged the difference between visual memory load and general cognitive (or attention) load on capture effects, we should also note that the interpretation that oculomotor capture is not purely stimulus-driven is in line with the recent study that investigated the effect of perceptual load on onset capture (Cosman & Vecera, 2009). Cosman and Vecera had observers search for a target letter through high-load and low-load displays in a variant of the flanker task (Lavie, 1995). Unlike a typical flanker paradigm, irrelevant flankers that included an onset and an offset appeared on each trial. If visual attention resources are limited (Lavie, 1995), increasing perceptual load on the search array should exhaust visual attention resources and result in modulation of onset capture. Cosman and Vecera found that onset flankers affected search in the low-load condition but not in the high-load condition. In line with the current study, Cosman and Vecera interpreted attenuation of onset capture in the high-load condition as evidence against the hypothesis that covert attention capture is purely stimulus-driven. Given both tasks that exhaust general cognitive and visual attention resources demonstrated attenuation of onset capture, it is possible that the observed modulation on oculomotor capture may not be modality specific (i.e., vision). However, until this non-modality specific account is directly tested, such an interpretation should be taken with caution.

In conclusion, we have presented initial evidence that oculomotor capture observed during real-world scene viewing is not purely driven by a bottom-up selection mechanism. Thus, oculomotor capture during scene viewing does not provide an example of automatic selection. Our results also have clear practical implications: Objects and events that may typically capture attention (e.g., a pedestrian stepping into a crosswalk) may fail to capture attention under higher cognitive load (e.g., a cell phone conversation). Additional research is necessary to determine the exact perceptual and cognitive processes that are involved in producing the observed interactions between bottom-up and top-down processes when attention and gaze are allocated to unexpected, unique and transient events in real-world scenes.

Acknowledgments

This research was made possible by a grant from the Economic and Social Research Council (RES-000-22-2484) awarded to James Brockmole and John Henderson. The experiments were conducted while Michi Matsukura (a postdoctoral fellow), James Brockmole, and John Henderson were at the University of Edinburgh. James Brockmole and John Henderson are currently Honorary Fellows of the University of Edinburgh. We thank Krista Ehinger for technical assistance and Stuart Ritchie for data collection assistance.

References

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46*, 2824–2833.

Boot, W. R., Brockmole, J. R., & Simons, D. J. (2005a). Attention capture is modulated in dual-task situations. *Psychonomic Bulletin & Review, 12*, 662–668.

Boot, W. R., Kramer, A. F., & Peterson, M. S. (2005b). Oculomotor consequences of abrupt object onsets and offsets: Onsets dominate oculomotor capture. *Perception and Psychophysics, 67*, 910–928.

Brockmole, J. R., & Henderson, J. M. (2005a). Prioritization of new objects in real-world scenes: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 857–868.

Brockmole, J. R., & Henderson, J. M. (2005b). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin & Review, 12*, 1061–1067.

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object-scene consistency. *Visual Cognition, 16*, 375–390.

Cosman, J. D., & Vecera, S. P. (2009). Perceptual load modulates attentional capture by abrupt onsets. *Psychonomic Bulletin & Review, 16*, 404–410.

Foulsham, T., & Underwood, G. (2007). Can the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123–1138.

Han, S. W., & Kim, M.-S. (2009). Do the contents of working memory capture attention? Yes, but cognitive control matters. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1292–1302.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498–504.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science, 16*, 219–222.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537–562). Oxford: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 5*, 438–443.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16*, 850–856.

Hunt, A., von Mühlenen, A., & Kingstone, A. (2007). The time course of attentional and oculomotor capture reveals a common cause. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 271–284.

Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research, 40*, 1443–1458.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13*, 201–214.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 451–468.

Lavie, N., & de Fockert, J. (2005). The role of working memory in attention capture. *Psychonomic Bulletin & Review, 12*, 669–674.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*, 363–386.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those fixations made during visual examination of briefly presented images. *Spatial Vision, 10*, 165–188.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11*, 157–178.

Matsukura, M., Brockmole, J. R., & Henderson, J. M. (2009). Overt attentional prioritization of new objects and feature changes during real-world scene viewing. *Visual Cognition, 17*, 835–855.

Olivers, C. N. L. (2009). What drives memory-driven attentional capture? *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1275–1291.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1243–1265.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107–123.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*, 125–154.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*, 3587–3596.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10*, 341–350.

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion pop-out phenomena. *Vision Research, 39*, 3157–3163.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 799–806.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9*, 379–385.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766–786.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43*, 333–346.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance, 33*, 363–377.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from selective search. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1505–1513.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.